ERDÖS' DISTINCT DISTANCES PROBLEM

Houssam Zenati
Centrale Supelec (Ecole Centrale) – France

Abstract. In discrete geometry, the Erdös' distinct distances problem states that between *n* distinct points in a plane there are at least $n^{1-o(1)}$ distinct distances. The problem was posed by Paul Erdös in 1946. In 2010, Larry Guth and Net Hawk Katz claimed to have a solution. The solution was published in 2015 in the *Annals of Mathematics*. This article aims at popularizing this problem to young students in mathematics, therefore no big background in mathematics is needed to understand it. It is open to every reader and shall be improved with any remarks or questions.
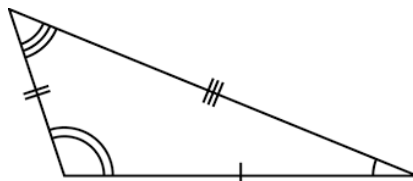*Keywords:* geometry; the Erdös' distinct distances problem; young students

I. Introduction
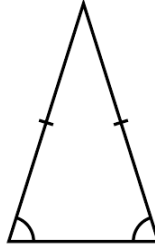Let's consider 3 points. We will consider the problem in a plane.

Figure 1. 3 points in a plane

Those 3 points constitute a triangle, and we can wonder how many distinct distances we can make with those 3 points. If the triangle has nothing special and is random, we can find 3 distinct distances.
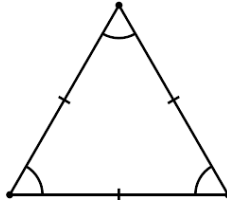

Figure 2. 3 distinct distances – random (arbitrary) triangle

However, if we consider an isosceles triangle, it has only 2 distinct distances.

Figure 3. 2 distinct distances – isosceles triangle

Besides, if we consider another special type of triangle, we can manage to have only one single distance for each point. This case corresponds to equilateral triangle.

Figure 4. One single distance between all points – equilateral triangle

In that configuration all distances have same value. The minimum number of **distinct** distances is 1.

Summary:
Given 3 points, the minimum number of distinct distances is 1.

II. Problem formulation and Overview
With this little overview of the problem considering 3 points, we can formulate the following general problem:
Given number *n,* what is the minimum number of distinct distances between *n* points in a plane?
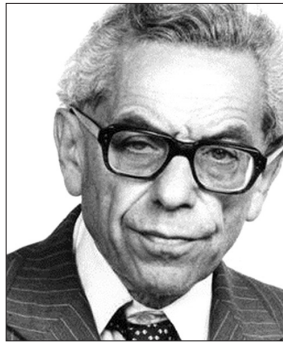***Definition**. $d(n)$* is the minimum number of distinct distances between *n* points.
We will consider the problem by dealing with the following points:
– Short biography
– Great dependency on initial points
– Combinatorial geometry
– Asymptotic approximations
The problem is about estimating the minimum number of distinct distances between *n* points, where *n* is a given number. With the example in (ii), we will see that

given a number n, we can have different positioning that give different number of distinct distances.

i – Short Biography[1)]

Figure 5. Paul Erdös

Paul Erdös (26 March 1913 / 20 September 1996) was a Hungarian mathematician. He was one of the most prolific mathematicians of the 20th century. He was known both for his social practice of mathematics (he engaged more than 500 collaborators) and for his eccentric lifestyle.

Erdös published around 1,500 mathematical papers during his lifetime, a figure that remains unsurpassed. He firmly believed mathematics to be a social activity, living an itinerant life style with the sole purpose of writing mathematical papers with other mathematicians. Erdös' prolific output with co-authors prompted the creation of the Erdös number, the shortest path between a mathematician and Erdös in terms of co-authorships. Erdös pursued and proposed problems in discrete mathematics, graph theory, number theory, mathematical analysis, approximation theory, set theory, and probability theory.
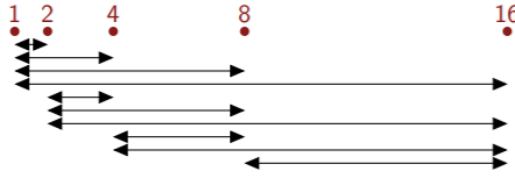
Ii – Initial Points Dependency

Considering nn points with coordinates $(k, 0)$ for k in $\{1, \ldots, n\}$:

Figure 6. Regular spacing between n points

The number of distinct distances is $n - 1$.. However, if we consider n points with coordinates $(2^k, 0)$ for k in $\{1, \ldots, n\}$

Figure 7. Regular spacing between n points

The number of distinct distances is $\frac{n(n-1)}{2}$. Indeed, we can prove that each distance is different from another. One can observe that these two examples show the dependency on initial points when considering the number of distinct distances. The problem will consist in determining the minimum number of distinct distances, therefore, we may consider changing positions for *n* given points.

Iii – Combinatorial Geometry

Combinatorial geometry is a blending of principles from the areas of combinatorics and geometry. It deals with combinations and arrangements of geometric objects and discrete properties of these objects. It is concerned with such topics as packing, covering, coloring, folding, symmetry, tiling, partitioning, decomposition, and illumination problems. Although combinatorial geometry was studied by classical mathematicians such as Euler and Kepler, many advances have been made since the middle of the 20th century. This topic was one which drew the interest of the late prolific mathematician Paul Erdös[1].

Iv – Asymptotic approximations

For small values of nn, it will be easy to determine the minimum number of different distances. But for larger values, we will need to use estimates of the value we are looking for. Therefore, we will define rapidly useful notations when doing asymptotic approximations. Those notations are popularized for a large public.

Let *g* and *f* be functions of *N* (i.e. $g(N)$ and $f(N)$ are two numbers which both depend on *N*).

- $g(N) = O(f(N))$ if and only if $|g(N)/f(N)|$ is bounded from above (i.e we can find *M* so that $|g(N)/f(N)|<M$ when *N* becomes "large enough". (We say that $N \to +\infty$)
- $g(N) \sim f(N)$ if and only if $|g(N)/f(N)|$ is as close as we want to 1, when *N* becomes "large enough". (We say that $N \to +\infty$)

Those notations will be used after we have seen first examples with small values of *n*, where *n* is the number of given points.

III. Examples for small values of *n*

Before trying to determine values for small values of nn, we can make the following observations:

● *n* is a positive integer
● For *n* points in a plane, we can find $\frac{n(n-1)}{2}$ possible distances. Indeed, it is easy to understand that for each point among *n* possible ones, it can be linked to another among *n* – 1 possible ones. But counting this way implies to count each distance twice. Therefore, the result has to be divided by 2. With reference to the problem with *d(n)*, where *d(n)* is the minimum number of distinct distances, we can observe that:

(1)
$$d(n) \leq \frac{n(n-1)}{2}$$

● We can also easily state that the minimum number of distances is at least 1, knowing that we will consider at least two points *n* ≥ 2, thus *d(n)* ≥ 1
● Last but not least, if we add one point to a disposition of *n* points, the minimum number of different distances is easily larger than the minimum number of different distances with *n* points:

(2)
$$d(n) \leq d(n+1)$$

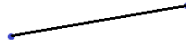**For** *n* = 2, there are only two points, therefore *d(2)* = 1.

Figure 8. 2 points

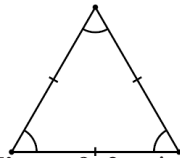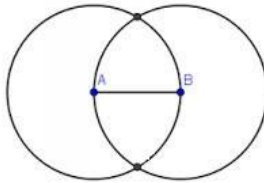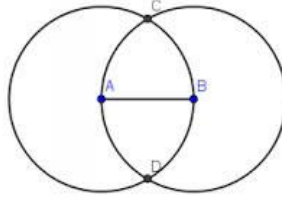**For** *n* = 3*n*, as we have seen, *d(3)* = 1,

Figure 9. 3 points

For *n* = 4, we know that *d(4)* ≥ 1. But is it possible to have *d(4)* = 1 ?
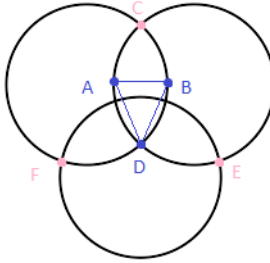
Proof Reduction Ad Absurdum. Let's suppose that *d(4)* = 1. Therefore, only one distance is available to set points. Let's sketch the figure. If we set two points, it will determine the only distance we can use when placing the other points.
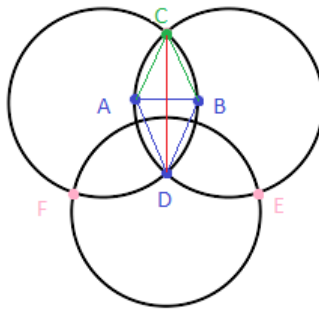
It is only possible to place the third point in one of those two intersections (i.e C or D), otherwise one distance between two points among three of those will be different.

Without lost of generality, let's place it in D, we then obtain the following figure (to obtain the other one, just reverse the figure).

Therefore, we have built the 3 first points, and still have to place the 4th one, i.e the last one. Since only one possible distance is to be used, the 4th point has to be at same distance to A and B, therefore has to be C. But if it is C, then the distance with D will be different (see the next figure).

Thus, this disposition is impossible, and C is not the 4-th point. We can use same reasoning for E and F, and therefore conclude that we cannot have only one possible distance to set 4 points at same distance to each other. We conclude that $d(4) = 1$ is absurd, and therefore $d(4) \geq 2$.
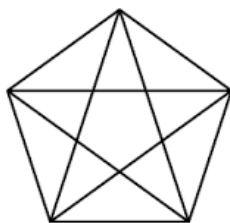
We can now find a positioning where we only use two different distances. This positioning is the one using a square, with one distance being the side of the square, and the other one being the diagonal of the square.

Figure 10. 4 points, using two different distances

Since we have $d(4) \geq 2$ with the previous proof and considering this positioning, we can therefore conclude that **for $n = 4$, $d(4) = 2$.**

For n = 5, we know that $d(5) \geq d(4)$ by using equation (2). Can we find a solution with 5 points using only 2 different distances?

Figure 11. 5 points, using two different distances

It is indeed possible with a regular pentagon. One distance is the side, the other one is the diagonal.

For $n = 6,7$ it is left as an exercise to the reader.

Hint: Prove that it is impossible to have $d(6) = 2$ and then find figures where $d(6) = 3$, $d(7) = 3$.

IV. Erdös Conjecture
The popularized principle of Erdös conjecture would be that the number $d(n)$ is "nearly close" to n when n grows.
We will try to popularize the mathematical formulation, though:

Conjecture. For any $a < 1$ we can find $C > 0$ for which $d(n) > C.n^a$ for n "big enough".$(n \to +\infty)$. Let's first explain what n^a means, where a is a real number.

i – Real number exponents
Let's first make a reminder.
Definition. Let n be a positive integer.
$a^n = a * a...*a$
  n times
$$a^{-n} = \frac{1}{a^n}$$
Then, we introduce the following definition.

Definition. Let p and q be positive integers.

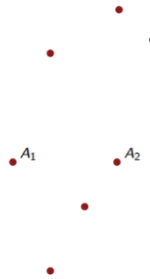$a^{\frac{p}{q}}$ is the number b for which we have:

$a^p = b^q$

It is therefore now comprehensible to $2,^{167}$ write as $1{,}67 = \frac{167}{100}$ and therefore we can write it $2^{\frac{167}{100}}$. We can also explain that rational numbers (i.e. those which can be written in the form $\frac{p}{q}$ where p is an integer and q is a strictly positive integer) are dense among real numbers, and we can "approximate" any real number by a rational one. We now understand the notation n^a in Erdös conjecture.

Ii – Proof for $\alpha = \frac{1}{2}$
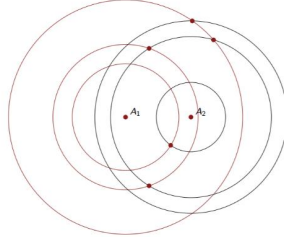Proof that $d(n) \geq C \sqrt{n}$. Let's consider nn points
$A_1, A_2,... A_n$ and we will focus on A_1 and A_2.

Figure 12. n points with A_1 and A_2

Let's consider circles with centers A_1 or A_2 passing through A_3, ..., A_n. Let C_1 denote the number of circles thus obtained centered at A_1 (i.e. redones) and C_2 the number of those centered at A_2 (i.e. black ones)

Figure 13. A_1 and A_2 with circles

We assume, without loss of generality (even if we change the names of the two points), that $C_1 \geq C_2$. Since two circles with different centers have at most two points of intersection, there are at most $2c_1c_2$ points of intersection in thefigure(c_1 choice for the first circle, c_2 choice for the second and at most 2 choices for theintersection). Now, by construction, each of the remaining $n - 2$points of our starting set belong both to a circle centered at A_1 and to a circle centered at A_2, so each of them is a point of the intersection between the two circles. We thus have $n - 2 \leq 2c_1c_2$ from which we deduce $n - 2 \leq 2c_1{}^2$. Therefore,

$$c_1 \geq \sqrt{\frac{n-2}{2}}$$

and we also know that $d(n) \geq c_1$. Thus,
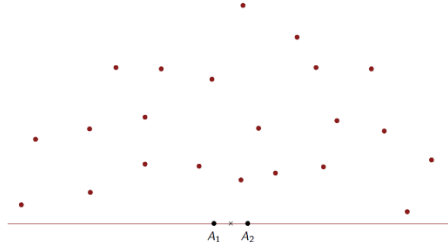
$$d(n) \geq \sqrt{\frac{n-2}{2}}$$

Using asymptotic approximations from II.iv, we can show easily that $n - 2 \sim n$, and therefore, when n is big enough, we eventually have:
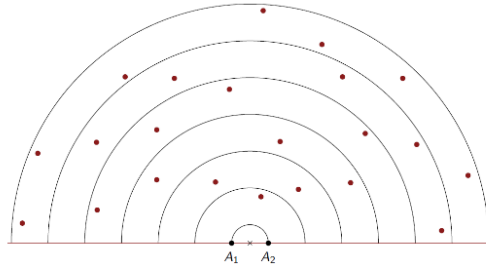
$$d(n) \geq C \sqrt{n}$$

Where $C = 1/\sqrt{2}$.

Iii – Proof for α = 2/3
Proof that $d(n) \geq C.n^{2/3}$. Let's consider n points $A_1, A_2, ..., A_n$ such that the minimum distance is reached between A_1 and A_2. The line (A_1, A_2) separates this cloud of points into two parts. One contains at least $\frac{n-2}{2}$ points, using the pigeon hole principle. From now on, it is limited to those points.
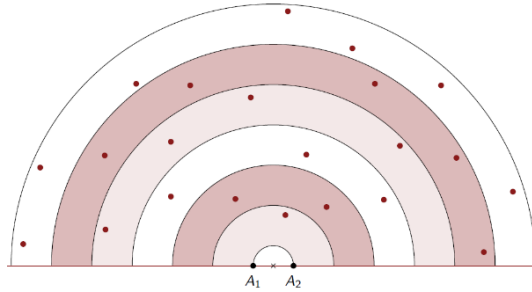
Figure 14. n points separated by (A_1, A_2)

We now sketch successive half-crowns centered on the midpoint of $[A_1 A_2]$ of width $A_1 A_2$ until we cover all the remaining points.

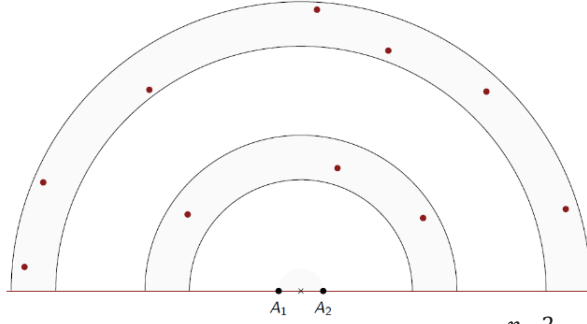**Figure 15.** n points separated by half-crowns

Let us divide these half-crowns modulo 3 into three sets.

Figure 16. Half-crowns modulo 3

At least one of these sets contains $\frac{n-2}{6}$ points, by using pigeonhole principle again. It is now limited to these points.

Note that points belonging to distinct half-crowns (among remaining ones) cannot be at the same distance of A_1 (respectively A_2).
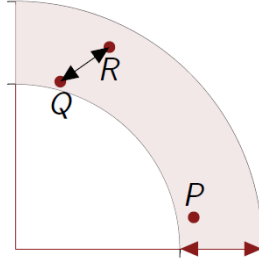
Figure 17. Half-crowns modulo 3, containing $\frac{n-2}{6}$ points

By using same reasoning than the first proof ($\alpha = 1/2$) we obtain that the n_i points of the half-crown i define at least $\sqrt{n_i}$ distances to A_1 or A_2. Thus, the number of distinct distances $\tilde{d}(n)$ for this configuration is greater or equal to (the sum relates to the half-crowns chosen) $\sum_i \sqrt{n_i}$. Therefore,

$$\frac{n-2}{6} \leq \sum_i n_i \leq \sqrt{\max(n_i)} \sum_i \sqrt{n_i} \leq \sqrt{\max(n_i)} \tilde{d}(n)$$

Now we consider the half-crown containing $\max(n_i)$ points, then half of the crown containing the majority of the points that is to say at least $\frac{1}{2}\max(n_i)$. Let P be a minimal ordinate point in this zone. A rapid calculation shows that if Q and R are equidistant from P in this half-crown, then $< 2A_1A_2$.

Figure 18. Quarter-crown containing $\frac{1}{2}$ max (n_i) points, with Q and P

There are thus at most two points of the zone at the same distance from P, which gives us at least $\frac{1}{2}\left(\frac{1}{2}\max(n_i) - 1\right)$ distinct distances. Hence, $max(n_i) \leq 4\,\tilde{d}(n) + 2$.

By deferring to previous inequality:

$$\frac{n-2}{6} \leq \sqrt{\max(n_i)}\,\tilde{d}(n)$$

we now obtain:

$$\frac{n-2}{6} \le \tilde{d}(n)\sqrt{4\tilde{d}(n)+2}$$

We now use asymptotic approximations knowing that $\tilde{d}(n) \to +\infty$, when n becomes big enough, the left term of the inequality is equivalent to n and the right term is equivalent to $2\tilde{d}(n)^{3/2}$, therefore, we have:

$$Cn^{2/3} \le \tilde{d}(n)$$

and eventually with general configuration:

$$d(n) \ge C.n^{2/3}.$$

where $C = 1/3^{\frac{2}{3}}$

V. Latest results

We will now give an overview of the latest results, with firstly a popularization of logarithm:

i – Logarithm

Logarithm is a function which takes only strictly positive real numbers as arguments.

– We remember that $2.2.2 = 8$, therefore $2^3 = 8$. The number 3 is actually called common logarithm of 8 in base 2.

– Likewise, $10.10.10.10 = 10000$

$10^4 = 10\ 000$

The number 4 is called common logarithm of 10000 in base 10.

– e is a special number, e = 2, 718...

– **Definition.** $ln(n)$ is number x for which $e^x = n$.

– $x = ln(n)$ may not be an integer, and can be negative

ii – Latest results

Notation. We write $f(n) = \Omega(g(n))$ when we can find $C > 0$ for which $f(n) > Cg(n)$ for n

"big enough".

Paul Erdös' 1946 lower bound of $d(n) = \Omega\left(n^{\frac{1}{2}}\right)$ was successively improved to:

- $d(n) = \Omega\left(n^{\frac{1}{3}}\right)$ (Leo Moser, 1952)

- $d(n) = \Omega\left(n^{\frac{5}{7}}\right)$ (Chung 1984)

- $d(n) = \Omega\left(n^{\frac{4}{5}}\ln(n)\right)$ (Chung, Szemerédi and Trotter 1992)

- $d(n) = \Omega\left(n^{\frac{4}{5}}\right)$ (Székely 1993)
- $d(n) = \Omega\left(n^{\frac{6}{7}}\right)$ (Solymosi and Tóth 2001)
- ...
- $d(n) = \Omega\left(\frac{n}{\ln(n)}\right)$ (Guth and Katz 2015)

Note that the latest result is interesting because it is commonly known that for any *a > 0* we have:

$$\frac{\ln(n)}{n^{\alpha}} \to 0$$

ln(*n*) is "weaker" than any n^a and we thus obtain the sharp exponent in the problem of Erdös.

NOTES

1. Paul Erdös Biography, Wikipedia.
2. Combinatorial Geometry, accessible at: http://mathworld.wolfram.com/CombinatorialGeometry.html

REFERENCES

Erdös, P. (1986). On some metric and combinatorial geometric problems. *Discrete Mathematics*, *60*, 147 – 153.

Mansuy, R. (2014). Les distances d' Erdös. *Images des Mathématiques*, CNRS.

Mansuy, R. (2013). *Le probléme des distances d'Erdös*.Erdös, P. (1946). On sets of distances of *n* points. *The American Mathematical Monthly, 53*, 248 – 250.

Guth, L. & N. H. Katz (2015). On the Erdös distinct distance problem in the plane. *ArXiV*.

ПРОБЛЕМЪТ НА ЕРДЬОШ
ЗА РАЗЛИЧНИТЕ РАЗСТОЯНИЯ

Резюме. В дискретната геометрия проблемът на Ердьош за различните разстояния твърди, че между *n* различни точки в равнината има поне $n^{1-o(1)}$ различни разстояния. Проблемът е поставен от Пол Ердьош през 1946 г. През 2010 г. Лари Тут и Нет Хоук Кац обявиха, че са го разрешили. Решението беше публикувано през 2015 г. в *Annals of Mathematics*. Целта на тази статия е да популяризира проблема сред младите ученици и студенти по математика, поради което не са необходими дълбоки математически познания за разбирането му. Статията е отворена за читателите и може да бъде подобрявана чрез забележки и въпроси.

✉ **Mr. Houssam Zenati**
Centrale Supelec (Ecole Centrale)
France
E-mail: houssam.zenati@centraliens.net